(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 164 556**
A2

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 85105405.6

(22) Date of filing: 03.05.85

(51) Int. Cl.⁴: C 12 N 15/00
C 12 N 1/16, C 12 P 21/02

(30) Priority: 11.05.84 US 609540

(43) Date of publication of application:
18.12.85 Bulletin 85/51

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: CHIRON CORPORATION
4560 Horton Street
Emeryville California 94608(US)

(72) Inventor: Rosenberg, Steven
3861 Fruitvale Avenue
Oakland California 94602(US)

(72) Inventor: Tekamp-Olson, Patricia
1426 43rd Avenue
San Francisco California 94122(US)

(74) Representative: Glawe, Delfs, Moll & Partner
Patentanwälte
Rothenbaumchaussee 58 Postfach 2570
D-2000 Hamburg 13(DE)

(54) Enhanced yeast transcription employing hybrid promoter region constructs.

(57) Yeast promoters of glycolytic enzymes are modified by isolating a fragment encompassing the RNA polymerase binding site and joining to the 5' end of this fragment a DNA sequence providing for enhanced inducible or constitutive transcription of a structural gene. Constructs are prepared for efficient expression of foreign genes in yeast.

Yeast strains 2150-2-3(pCl/1GAPSOD) and AB110(pCl/1GAPATI9), producing human α_1-antitrypsin and superoxide dismutase, were deposited at the A.T.C.C. on May 9, 1984 and given Accession Nos. 20708 and 20709, respectively; and 2150-2-3(GAP5), 2150-2-3(Pyk5) and 2150-2-3(PHO5GAP1), expressing Hepatitis B surface antigen, were deposited at the A.T.C.C. on May 9, 1984 and given Accession Nos. 20705, 20706 and 20707, respectively.

EP 0 164 556 A2

Croydon Printing Company Ltd

ENHANCED YEAST TRANSCRIPTION EMPLOYING HYBRID PROMOTER REGION CONSTRUCTS

BACKGROUND OF THE INVENTION

Field of the Invention

With the advent of hybrid DNA technology, production of mammalian proteins in microorganisms became a reality. For the first time, mammalian proteins could be synthesized in a unicellular microorganism by introduction of a gene encoding for the mammalian protein under the transcriptional and translational control of regulatory sequences recognized by the microorganism host. The introduction of these foreign constructions into a microorganism host resulted in competition between the regulatory signals of the construct and the regulatory signal endogenous to the host for the host systems involved with expression. The structural gene of interest is usually directed to a product which is nonproductive and may be injurious to the host. Therefore, host cells which can turn off the foreign gene can effectively dominate modified host cells.

Substantial progress has been made in isolating sequences concerned with transcriptional and translational regulation for protein expression. But frequently flanking sequences, as well as distant sequences, may also affect the efficiency and regulation of the expression of the protein. Therefore, as one manipulates these various sequences, removing them from their native environment, and joining them to unnatural sequences, that is sequences other than the wild-type sequence, one can never be certain as to the result.

In order to enhance the economies of producing proteins in microorganisms, there have been substantial efforts directed to improving the efficiency of transcription and translation, maximizing the proportion of

total protein directed to production of the desired product, enhancing the viability of the modified host, as well as improving the efficiency with which the modified host may be obtained.

Description of the Prior Art

Guarente et al., Proc. Natl. Acad. Sci. USA (1982) 79:7410-7414, describes a hybrid promoter region employing the GAL4 regulatory region. Guarente and Ptashne, ibid. (1981) 78:2199-2203, report the presence of two domains in a yeast promoter, with a region upstream from the TATA sequence providing an activation site. Kramer et al., ibid. (1984) 81:367-370, describe the regulated expression of a human interferon gene in yeast employing the yeast acid phosphatase promoter, where expression is induced by phosphate or a temperature shift. Tekamp-Olson et al., Cold Spring Harbor Meeting, Molecular Biology of Yeast, 1983, describe the absence of deleterious effects on yeast viability when employing "short" promoters, as distinct from the presence of such effects, when employing an extended promoter region or "long" promoters.

SUMMARY OF THE INVENTION

Novel hybrid promoter regions are provided for use in conjunction with constructs having a structural gene under the transcriptional control of the hybrid promoter region and a terminator region. The hybrid promoter comprises a first segment providing transcriptional enhancement, either constitutive or regulated, and a second segment defining an RNA polymerase binding site and transcriptional initiation site. The hybrid promoter regions lack the deleterious effects associated with a wild-type promoter region in recombinant constructs, which results in reduced transformation efficiencies and lower yeast viability.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagrammatic view of plasmids GAP1-6;

Figure 2 is a diagrammatic view of plasmids Pyk1-6; and

Figure 3 indicates the DNA linker sequence and a flow diagram showing its use in a construct for hSOD.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Novel DNA fragments are provided as well as constructions employing the fragments for enhanced expression of structural genes in a yeast host. The constructs employing a hybrid promoter region provide for enhanced efficiencies of transformation and greatly improved viability of the yeast host as contrasted with those employing a wild-type yeast promoter. Concomitant with the improved viability is increased expression of a structural gene, in comparison with the truncated promoter region, and, therefore, greatly enhanced overall yields of expression products.

For the purposes of the subject invention, the "promoter region" is divided into two domains: (1) the structural gene proximal region, which includes the transcription initiation site, the "TATA" sequence capping sequence, as appropriate, and an RNA polymerase binding sequence, which sequence intends a sequence which includes nucleotides upstream from the initiation site for directing the initiation of synthesis of the messenger RNA; and (2) a distal region which provides for regulated or constitutive expression, with enhanced expression as compared to the first domain linked to non-functional yeast DNA.

The hybrid promoters of the subject invention employ the RNA polymerase binding region of a yeast glycolytic enzyme promoter and a region upstream from said polymerase binding region, which is different from

the wild-type upstream region of the RNA polymerase binding region and provides for enhanced efficiencies of transcription. This distal region will be derived from either a sequence, usually a yeast sequence, involved in regulation of transcription, or a prokaryotic sequence which provides for enhanced constitutive expression of the desired gene.

Conveniently, cassettes or constructs can be prepared which provide for one or more restriction sites intermediate the promoter region and a related terminator region where the structural gene may be inserted, so as to be under the transcriptional control of the hybrid promoter region. By having one or more restriction sites, one can provide for ease of insertion of the structural gene intermediate the transcription initiation and termination regions. The cassettes which can be prepared comprising the transcriptional initiation and termination region, having appropriate restriction sites for structural gene insertion can be cloned in prokaryotic vectors, so that after insertion of the structural gene, the resulting cassette, including the structural gene, may be cloned, isolated and purified, before introduction into a yeast vector.

The cassette, will for the most part, have the following formula:

-(P.R.(2)-P.R.(1))-R.S.-T.R.-

wherein:

P.R.(1) is the promoter region proximal to the structural gene and having the transcription initiation site, the RNA polymerase binding site, and including the TATA box, the CAAT sequence, as well as translational regulatory signals, e.g., capping sequence, as appropriate;

P.R.(2) is the promoter region joined to the 5'-end of P.R.(1) associated with enhancing the

efficiency of transcription of the RNA polymerase binding region;

R.S. is a sequence having one or more restriction recognition sites, preferably at least two restriction recognition sites, where the sites may result upon restriction into blunt ends or overhangs;

T.R. intends the termination region, which will include the terminator, which may be a stem and loop structure, and which may be associated with one or more stop codons, a polyadenylation signal sequence, if any, as well as any other transcriptional and translational termination sequences.

P.R.(1) will generally be at least about 150bp, more usually at least about 200bp, usually not more than about 600bp, more usually not more than about 500bp, generally not more than about 450bp and preferably less than about 400bp; the sequence will extend in the downstream direction of transcription to about bp +3, more usually bp -1 and may extend only to bp -20, more usually to bp -10 (the numbering intends that +1 is the first bp of the initiation codon with which the promoter region is associated in the wild-type host, while -1 is the immediately upstream bp and the integers increase in the direction of transcription;

P.R.(1) will be derived from a strong yeast promoter, normally a glycolytic enzyme promoter, such as glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, alcohol dehydrogenase, phosphoglucoisomerase, triose phosphate isomerase, phosphofructokinase, etc.;

P.R.(2) will provide for an enhancing function of transcription, which enhancing function may provide for constitutive or regulated transcription; regulators will be derived from regions associated with regulation of yeast genes, other than the natural or wild-type gene associated with the first domain in the wild-type or natural host, such as UDP-galactose epimerase (GAL10), galactokinase (GAL1), acid phosphatase (PHO5), alcohol

dehydrogenase I and II (ADH I and II), etc. For yeast regulatory sequences, the domain will usually be at least about 100bp, more usually at least about 200bp, for convenience generally not exceeding about 3kbp, usually not exceeding about 1kbp, desirably not exceeding about 600bp. The regulatory region will generally begin at least about 200bp from the initiation codon, usually at least about 300bp and may begin at 400bp or farther upstream from the initiation codon.

Regulation can be as a result of a change in the chemical or physical environment of the host, such as a change in carbon source, e.g., glucose to galactose or vice versa; a change in concentration of a nutrient, e.g., an inorganic nutrient such as a phosphate; or a change in temperature, e.g., 25°C to 35°C. Constitutive transcription can be achieved employing prokaryotic sequences of at least about 500bp, usually 1kbp or more, for convenience, generally not exceeding about 5kbp; conveniently, the prokaryotic sequence can be obtained from the vector in which the cassette is cloned, illustrative vectors including pBR322, lambda, Charon 4A, pACYC184, pUC5, etc.

R.S. will generally be at least 4bp, more usually at least 6bp, and may be 100bp or more, more usually being not more than about 60bp and may include one or more, usually not more than about 10 restriction sites, where such restriction sites may be illustrated by EcoRI, BamHI, SalI, HindIII, AluI, AvaI, TaqI, HpaI, etc., having at least one unique restriction site for the construct sequences.

T.R. is the termination region which will include the necessary transcriptional and translational signals for termination, such as the polyadenylation site, etc.;

T.R. will generally be at least about 100bp, more usually at least 150bp, and generally less than

about 1kbp, usually less than about 600kbp; the termination region may be derived from any convenient yeast sequence, so long as the terminator balances the promoter, conveniently being derived from a glycolytic enzyme terminator, where the terminator may be associated with the same or different enzyme with which the promoter is associated.

Where a cassette is cloned in a bacterial vector, the construction will have the following formula:

$$-(P.R.(2)-P.R.(1))-R.S.-T.R.-Rep(B)-(M(B))_a-\Big|_{b}(-)$$

wherein all the symbols have been defined previously, except for:

Rep (B), which intends a replicon or replication system recognized by a prokaryotic host and may be derived from a plasmid or phage, such as ColE1, and R plasmid, e.g., pRK290, lambda, e.g., Charon 4A, λdv, etc.;

M is a marker which provides for selection of hosts containing the construction, where (B) intends a prokaryotic, e.g., bacterial, host and a intends an integer of from 0 to 3, usually 1 to 2, although additional markers may be present, where the marker allows for selection of the host containing the construct as well as providing for selective pressure on maintaining hosts having the construct; the markers include biocide resistance, such as antibiotic resistance, toxin resistance and heavy metal resistance; providing prototrophy to an auxotrophic host; providing immunity; and the like;

the markers may provide for complementation of an auxotrophic host, e.g., his^-, ura^-, trp^-, leu^- genotype, resulting in prototrophy; resistance to metals,

such as cup^+ genotype; resistance to antibiotics, such as amp^r, tc^r, cam^r, str^r, tur^r genotype, etc.;

b is 0 or 1, intending that the construction is either linear or circular, usually circular.

The above construct can be used for insertion of a wide variety of structural genes, both prokaryotic and eukaryotic, both naturally occurring and synthetic, where the genes may include signal leaders for secretion, and the like. The genes may express enzymes, hormones, proteins from pathogens for vaccines, structural proteins, lymphokines, membrane surface proteins, immunoglobulins, blood proteins, or the like. The particular structural gene which is inserted is not critical to this invention and any polypeptide or protein of interest may be prepared employing the constructions of the subject invention. The structural genes will usually be foreign to the yeast host, where foreign intends different from wild-type yeast structural genes and from a source that does not normally exchange genetic information with yeast.

Usually, the structural gene will be at least about 36bp, and not more than about 20kbp, usually not more than about 3000bp, usually not more than about 1500bp. Included in the structural gene may be non-coding flanking regions, the 5'-flanking region normally being quite short, usually less than about 30bp, while the 3'-flanking region may be extended, usually not exceeding about 500bp. Thus, the structural gene fragment will usually include the translational stop codons for proper termination of amino acid chain extension.

When the structural gene has been inserted into the cassette which is joined to a yeast replication system, normally including one or more markers recognized by yeast, the resulting construct will have the following formula:

$$-(P.R.(2)-P.R.(1))-gene-T.R.-Rep(Y)-(M(Y))_a-$$

wherein all of the symbols have been defined previously except for:

gene, which intends the structural gene, having its initiation codon and stop codons as appropriate; and

Y, which intends that the symbol is related to yeast.

Convenient yeast replication systems include the 2μm plasmid replication system, combination of CEN3 and ARS1 or ARS3, or the like. The replication systems may be high or low copy number, depending on the effect of the construct on the viability of the host. While the indicated replication systems are those which have found common employment, any replication system useful in yeast may be employed which provides for efficient replication and maintenance. Often the structural gene will be inserted into an appropriate shuttle vector capable of replication and selection in either a yeast or bacterial host, where the resulting construction will have the following formula:

$$-(P.R.(2)-P.R.(1))-gene-T.R.-Rep(Y)-(M(Y))_a-Rep(B)-(M(B))_a-$$

where all symbols have been defined previously. Also, it is, of course, understood that the cassette without an inserted structural gene but containing the restriction enzyme recognition sequence, R.S., may be propagated in yeast or contained within a shuttle vector, where the construction will have the following respective formulae:

$$-(P.R.(2)-P.R.(1))-R.S.-T.R.-Rep(Y)-(M(Y))_a-$$

$$-(P.R.(2)-P.R.(1))-\ R.S.-T.R.-Rep(Y)-(M(Y))_a-Rep(B)-(M(B))_a-$$

where all symbols have been defined previously.

The various fragments which form the cassette and final constructions may be joined together in accordance with conventional ways. In many cases, genes have been isolated and restriction mapped, as well as sequenced. To that extent, one can select the sequence of interest by restriction of the gene, employing further manipulation as necessary such as resection with Bal31, in vitro mutagenesis, primer repair, or the like, to provide a fragment of a desired size, including the desired sequence, and having the appropriate termini. Linkers and adapters can be used for joining sequences, as well as replacing lost sequences, where the restriction site is internal to the region of interest. The various fragments which are isolated, may be purified by electrophoresis, electroeluted, ligated to other sequences, cloned, reisolated and further manipulated.

The use of regulatory sequences for controlling transcription of the structural gene of interest allows for growing the host cells to high density with no or low levels of expression of the structural gene, and then inducing expression by changing the environmental conditions, e.g., nutrient, temperature, etc.

For example, with the GAL4 regulatory region, the yeast cells could be grown in rich media with a glycerol-lactic acid combination to high density, e.g., mid or late log phase, followed by switching the carbon source to galactose. For PHO5 regulation one could grow the cells at high phosphate, about 1 to 10mM, and then decrease the phosphate concentration to about 0.1 to 0.5mM. For temperature sensitivity, one could grow the cells at 25° to 37°C and then change the temperature as appropriate by about 5° to 20°C. The host cells would have the regulatory system associated with the regulatory region employed.

Various techniques will be exemplified in the Experimental section of this application, which techniques can be used as paradigmatic for constructions employing fragments from sources other than those exemplified. Of particular interest, as evidenced by the Experimental section, will be the use of the glyceraldehyde-3-phosphate dehydrogenase promoter region for the RNA polymerase binding site in conjunction with regulator sequences, such as those associated with GAL4, PHO5, ADHII or the like. In referring to the GAL4 regulatory region or associated sequence, the region intends the sequence associated with regulation of other galactose metabolism genes, e.g., GAL1 and GAL10, which are under the regulatory control of such sequence in conjunction with the expression product of the GAL4 gene. The PHO5 and ADHII sequences refer to regions associated with the PHO5 and ADHII genes which provide for transcriptional regulation of the PHO5 and ADHII gene, respectively.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

All DNA manipulations were done according to standard procedures. See Molecular Cloning, T. Maniatis et al., Cold Spring Harbor Lab., 1982. Enzymes used in cloning were utilized as per the manufacturer's specifications. Enzymes were obtained either from New England Biolabs or Bethesda Research Laboratories. Procedures with these enzymes employed the supplier's directions. Yeast were transformed and grown using a variety of media including selective medium (yeast nitrogen base without leucine); YEPD medium, containing 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) glucose, and others as appropriate and/or detailed below. In the case of plating medium contained 2% (w/v) agar and for transformation 3% top agar. Hepatitis B surface antigen

was determined after lysis of yeast by glass bead agitation and clarification by centrifugation, using the AusriaII assay (Abbott Laboratories). Protein is determined by the Coomassie dye binding method.

Construction of GAL regulator containing plasmids.

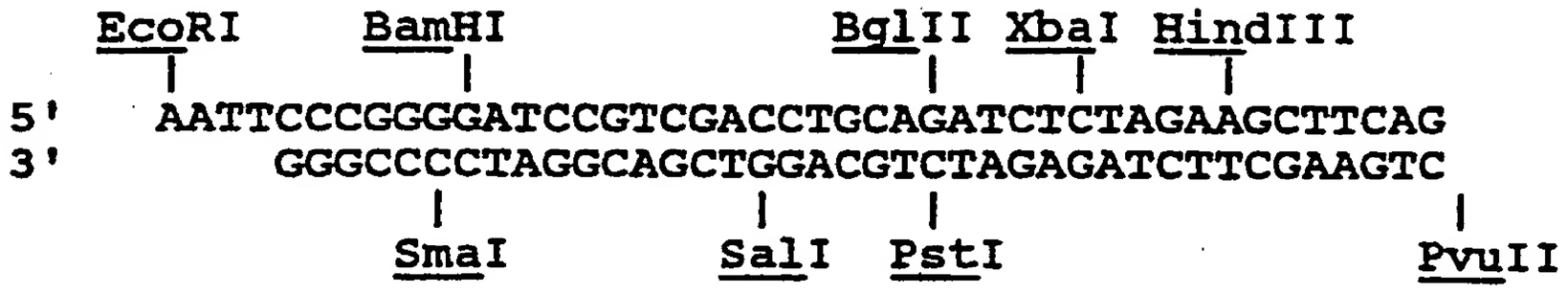
Plasmid pLGSD5 is prepared as described in Guarente et al., (1982) supra. The plasmid was manipulated as follows: After restriction with XhoI, the overhangs were filled in with the Klenow fragment of DNA polymerase I ("Klenow fragment"), ligated with EcoRI linkers (GGAATTCC) and then completely digested with EcoRI and Sau3A to provide a 370bp fragment which was isolated by gel electrophoresis and included the intergenic sequence between GAL1 and GAL10 genes of yeast, and provides for the GAL4 regulation sequence of the GAL1 and GAL10 genes.

This fragment was inserted into pBR322 which had been completely digested with EcoRI and BamHI, followed by treatment with alkaline phosphatase to prevent oligomerization. The resulting plasmid pBRGAL4 was treated in two different ways.

In the first procedure pBRGAL4 was completely digested with Sau3A, the overhangs filled in with the Klenow fragment, and the resulting blunt-ended fragment ligated with SalI linkers (CGTCGACG), followed by digestion with SalI and XhoI. The resulting 370bp fragment was isolated by gel electrophoresis. This fragment has the original 370bp yeast GAL4 regulator sequence with XhoI and SalI termini.

The second fragment was obtained by complete digestion of pBRGAL4 with XhoI and SalI to provide a XhoI-SalI fragment which included the 370bp yeast GAL4 regulator sequence as well as about 280bp of pBR322, the GAL4 sequence extending from Sau3A to SalI.

The two fragments were then cloned in the plasmid plot5. plot5 was prepared by inserting the 40bp polylinker fragment of the following sequence



into pBR322 as an EcoRI-PvuII substitution followed by insertion of the trp-lac promoter (Russell and Bennett, Gene (1982) 20:231-245) into the PvuII site with transcription oriented toward the polylinker sequence. plot5 was completely digested with SalI, followed by treatment with alkaline phosphatase and the 370bp and 650bp fragments independently inserted into plot5 to provide plasmids plot5GAL4/370 and plot5GAL4/650, respectively. Each of the plasmids was then completely digested with BamHI and SalI to reproduce the individual fragments extended by 6bp of the polylinker fragment. These fragments were then ligated into pC1/1, which had been completely digested with BamHI and SalI followed by treatment with alkaline phosphatase to prevent recircularization. Plasmid pC1/1 is a derivative of pJDB219 (Beggs, Nature (1978) 275:104) in which the region corresponding to bacterial plasmid pMB9 in pJDB219 has been replaced by pBR322 in pC1/1. The resulting plasmids were designated pC1/1GAL4/370 and pC1/1GAL4/650, respectively. The BamHI-SalI fragment is located in the pBR322 portion of the vector pC1/1.

The next construction develops a hybrid promoter for expression of the Hepatitis B surface antigen (HBsAg or sAg), employing the RNA polymerase binding region of GAPDH. The previously prepared plasmid

pHBS56/16-3, a yeast shuttle vector containing the alcohol dehydrogenase 1 (ADH1) promoter, the HBsAg gene and ADH terminator as a SphI fragment, was digested with SphI and the ends modified with Bam linkers. The Bam linkers have the sequence CGGATCCG.

pHBS56/16-3 was prepared as follows: A TaqI-HpaI fragment obtained from the HBsAg coding region which included 26bp of the pre-sAg region, 681bp of the sAg region and 128bp of the 3'-untranslated region, was linked with EcoRI linkers and cloned at the EcoRI site in pBR322. The EcoRI linkers have the sequence GGAATTCC. The plasmid pHBS5 was thus obtained.

After digesting pHBS5 with EcoRI, the digest was resected with Bal31 and religated with EcoRI linkers (GGAATTCC). After digestion with EcoRI the material of about 800bp was isolated from a polyacrylamide gel. This isolate was then recloned into pBR322 which had been digested with EcoRI and treated with alkaline phosphatase. Where the resection was to the sequence CATGG, which included the methionine codon, the EcoRI linker created an NcoI site. The plasmids were screened for the presence of an NcoI site and one of the plasmids chosen for further manipulation. This plasmid, designated pHBS5-3, was restricted with EcoRI, the EcoRI fragment made blunt-ended with Klenow fragment and dNTPs, and the blunt-ended fragment was then restricted with XbaI to provide an about 100bp fragment having an XbaI overhang and blunt end at the former EcoRI site.

pHBS5 was then digested with ClaI, made blunt-ended with the Klenow fragment and dNTPs, digested with XbaI, followed by alkaline phosphatase treatment. The 100bp fragment was then inserted into the vector to provide the plasmid pHBS6. Upon sequencing of the blunt-ended ligation site, it was found that an adenosine had been lost, so as to lose the EcoRI site, where the sequence was now ATCGATTCCCATGG. The ClaI and NcoI

sites were retained. The loss of the A resulted in pHBS6 having a single EcoRI site.

pHBS5-3 was digested with EcoRI and the resulting EcoRI fragment having the sAg fragment isolated by gel electrophoresis and inserted into the vector pHBS16 (Valenzuela et al., Nature (1982) 298:347-350). This plasmid has the ADH1 promoter and the sAg gene in an EcoRI fragment in a plasmid containing the 2μm origin, a TrpI gene and pBR322. The plasmid was digested with EcoRI, treated with alkaline phosphatase to prevent recircularization, and the EcoRI fragment from pHBS5-3 inserted to provide pHBS16-3, where the sAg gene isolated as a TaqI-HpaI fragment had been modified by Bal31 resection. The plasmid pHBS16-3 was digested with SphI and XbaI to provide a fragment which had the ADH promoter at the Sph terminus and the 5'-end of the sAg gene.

pHBS56 was then digested with SphI. pHBS56 was prepared from pC1/1 by digestion with SphI, which deletes a portion of the plasmid spanning the 2μm-pBR322 joint. The active portion of the ADH1 promoter region is contained within the SphI-HindIII fragment of approximately 300bp (Bennetzen et al., J. Biol. Chem. (1982) 257:301). The SphI site in the ADH promoter begins at position -413 and the yeast terminator sequence is contained within a HindIII-SphI fragment of about 330bp. In each case the SphI site is distal to the coding region. A 1500bp ADH1 promoter fragment terminating at position -9 (Hitzeman et al., Nature (1981) 293:717) and an approximately 450bp terminator unit from nucleotides 913 to 1368 in the ADH gene nucleotide sequence were joined at a HindIII site between the fragments and cloned into the BamHI site of the vector YEp13 (Broach and Hicks, Gene (1979) 8:121) to provide pADH5.

The HBsAg-DNA segment of pHBS5 was excised by EcoRI digestion, blunt-ended with the Klenow fragment and joined at both ends with HindIII linkers, CAAGCTTG. After digestion with HindIII, the HBsAg fragment was

inserted into the HindIII site of the plasmid pADH5 which had been digested at the HindIII site intermediate the ADH1 promoter and terminator sequence. A plasmid with the HBsAg gene in the correct orientation as determined by restriction analysis was designated pHBS22. The cassette was included between two SphI restriction sites. pHBS22 was digested with SphI to obtain a fragment of about 1500bp and inserted into SphI digested pC1/1 to provide pHBS56 which was cloned in E. coli HB101.

pHBS56 was digested with SphI and XbaI to provide a 1.1kb fragment having the ADH terminator region and the 3'-portion of the sAg gene with the SphI site proximal to the terminator region. The 1.1kb SphI-XbaI fragment was joined to the SphI-XbaI fragment from pHBS16-3, which resulted in providing the complete sAg gene in the correct orientation between the ADH promoter and terminator. This SphI-SphI fragment was then ligated to SphI digested pHBS56, replacing the cassette of pHBS56 to provide the plasmid pHBS56/16-3 with the resected sAg coding region fragment. The cassette was then excised from pHBS56/16-3 by digestion with SphI, followed by chewing back the overhangs with the Klenow fragment in the presence of dNTPs, then ligated with BamHI linkers, followed by digestion with BamHI to provide a 1.6kb fragment which was isolated by gel electrophoresis. The fragment included the ADH promoter region, the sAg gene and ADH terminator region, as described above. This fragment was inserted into the BamHI site of pBR322 to provide pPGT16-3 which was digested with BamHI and XbaI and the resulting 1.1kb fragment gel isolated, where the XbaI-BamHI fragment had the 3' portion of the sAg gene and the ADH terminator region.

pHBS6 was digested with XbaI and NcoI and the 94bp fragment gel isolated to provide the 5'-portion of

the sAg gene. A synthetic adapter was prepared of the formula

$$CGA_2TA_3(CA)_3TA_3CA_3CAA$$
$$T_2AT_3(GT)_3AT_3GT_3GTTGTAC$$

having TaqI and NcoI termini and providing the -25 to -1 nucleotides of the GAPDH (GAP49) promoter and the initiation codon of the sAg gene. This synthetic fragment, the NcoI-XbaI fragment, and the XbaI-BamHI fragment were ligated simultaneously, followed by digestion with TaqI and BamHI. The resulting fragment was then substituted into pBR322 linearized with ClaI and BamHI, followed by treatment with alkaline phosphatase. The resulting plasmid, which contains the -1 to -25bp of the GAPDH promoter region, the sAg gene, and the ADH terminator, where the NcoI restriction site is lost was called pHBS6LGAPsAgtADH.

pGAP1, a plasmid prepared by insertion of a HindIII fragment containing the GAPDH gene GAP49 (Holland and Holland, J. Biol. Chem. (1979) 254:5466-5474) inserted in the HindIII site of pBR322, was digested with HinfI and a 500bp promoter containing fragment isolated. The fragment was resected with Bal31 to remove about 50 or 90bp, followed by ligation with HindIII linkers and digestion with HindIII. pBR322 was digested with HindIII, followed by treatment with alkaline phosphatase and the about 450 or 410bp fragment inserted to provide pGAP128 and pGAP396, respectively.

pGAP128 was digested with HindIII, the fragment made blunt-ended with the Klenow fragment and dNTPs and the resulting 450bp fragment isolated by gel electrophoresis. This fragment was inserted into SmaI digested plot5, which had been treated with alkaline phosphatase, to provide plasmid plot5pGAP128, which contained about -400 to +27bp of the GAPDH promoter and coding region. Plasmid plot5pGAP396 was prepared from pGAP396 in an identical manner and thus differs from plasmid

plot5pGAP128 in having about 15-30 fewer bp at each terminus of the GAPDH promoter region (about -385 to -3).

Plasmids GAP1-GAP4 were then prepared in the following manner. Plasmid plot5pGAP128 was digested with TaqI and BamHI to provide an about 390bp TaqI-BamHI fragment which included the -26 to about -400bp of the GAPDH promoter region and a portion of the HindIII and plot5 polylinker. pHBS6LGAPsAgtADH plasmid was also digested with TaqI and BamHI and a 1.1kb TaqI-BamHI fragment containing the 3'-terminus of the GAPDH promoter region, the sAg gene and the ADH terminator region was gel isolated and ligated to the other TaqI-BamHI fragment to provide a BamHI-BamHI fragment which included approximately 400bp of the GAPDH promoter region, the sAg gene in proper orientation for transcriptional regulation by the GAPDH promoter, followed by the ADH terminator region. This fragment was ligated into pBR322 which had been digested with BamHI and treated with alkaline phosphatase to provide plasmid pPGT80. This BamHI cassette could now be isolated and inserted into plasmid pC1/1, at the BamHI site in the pBR322 portion of pC1/1, where in plasmid GAP1 the ADH terminator region is proximal to the amp^r gene with the pBR322 portion divided into an approximately 4kb sequence including the amp^r gene and a 375bp region separating the cassette from the 2μm sequences. In GAP2, the promoter is adjacent to the long pBR322 sequence with transcription in the same direction as the amp^r gene. The same cassette was inserted into BamHI-digested pC1/1GAL4/650 to obtain plasmids GAP3 and GAP4, where GAP3 has the GAPDH promoter distal from the GAL4 regulator region and the long pBR322 sequence and GAP4 has the GAPDH promoter adjacent to the GAL4 regulator region, which is adjacent to the long pBR322 sequence.

Plasmids GAP5 and GAP6 were isolated as follows. Plasmid plot5pGAP396 was digested with SalI and TaqI

and a fragment containing 9bp of the plot5 polylinker sequence and the GAPDH promoter sequence extending from about -385 to -26bp was isolated. An approximately 130bp TagI-XbaI fragment including -25 to -1bp of the GAPDH promoter and +1 to +93bp of the sAg gene was obtained from pHBS6LGAPsAgtADH. A 1.1kb XbaI-SalI fragment containing the 3'-portion of the sAg gene and the ADH terminator as well as 6bp of plot5 polylinker sequence was obtained from plasmid plot5sAgtADH (described below--Pyruvate Kinase Promoter). These three fragments were ligated, digested with SalI and then cloned into SalI-digested pC1/1GAL4/370. GAP5 has the GAPDH promoter region adjacent to the GAL4 regulator region, which is proximal to the short pBR322 sequence, and GAP6 has the GAPDH promoter region distal from the GAL4 regulator region and proximal to the long pBR322 sequence (see Fig. 1).

Pyruvate kinase promoter.

Plasmid pHBS6Pyk containing the sAg gene under the transcriptional regulatory control of the Pyk promoter was obtained by cloning a 4.4kb insert of yeast genomic DNA in pBR322 containing the Pyk gene and 911 nucleotides of 5'-untranslated region, and digestion of this plasmid pPyk9.1.1 with XbaI. After making the ends blunted-ended, the linear fragment was digested with BamHI providing a 912bp BamHI-blunt fragment containing the Pyk promoter and 8 bases from the Pyk coding region. This fragment was inserted into the plasmid pHBS6, which had been digested with NcoI, blunt-ended and digested with BamHI. The plasmid pHBS6Pyk was totally digested with EcoRI, to obtain a fragment including the sAg gene and a portion of the Pyk promoter region. The fragment was made blunt-ended with the Klenow fragment and dNTPs, followed by ligation to BamHI linkers, digested with XbaI, which is internal to the sAg gene, the XbaI terminus made blunt-ended with the Klenow

fragment and dNTPs, followed by digestion with BamHI, to provide a 580bp BamHI-blunt-ended (XbaI) fragment. The plasmid plot5 was digested with EcoRI, made blunt-ended, digested with BamHI and treated with alkaline phosphatase and the two fragments joined to provide plasmids plot5PyksAg51 and plot5PyksAg.57. The two differ in that the BamHI site of the latter was not regenerated during cloning, possibly as a consequence of minimal nuclease contamination (digestion).

plot5 was treated as previously described (EcoRI digestion, blunt-ended, BamHI digestion and treatment with alkaline phosphatase) and joined to a 1.1kb fragment obtained by digestion of pPGT16-3 with XbaI, followed by blunt ending, followed by digestion with BamHI and gel isolation. This fragment was introduced into plot5 to provide the plasmid plot5sAgtADH. Again the BamHI site in this plasmid was not regenerated, presumably due to digestion by contaminating nuclease.

Plasmids Pyk1 and Pyk2 were prepared as follows. Plasmid plot5PyksAg51 was digested with BamHI, then with XbaI, and an approximately 580bp fragment containing about 480bp of Pyk promoter and 93bp of the 5'-end of the sAg gene was gel isolated. A 1.1kb XbaI-SalI fragment containing the 3'-portion of the sAg gene, the ADH terminator and about 6bp of the plot5 polylinker was isolated from plot5AgtADH. These two fragments were ligated, digested with BamHI and SalI and then cloned into plasmid pC1/1, which had been cleaved with BamHI and SalI and treated with alkaline phosphatase, to yield plasmid Pyk1. Plasmid Pyk2 was prepared similarly but the 580bp SalI-XbaI, Pyk promoter/HBsAg gene 5'-end fusion fragment was isolated from plot5PyksAg.57 and included about 6bp of plot5 polylinker sequence upstream from the promoter region. Also the 1.1kb XbaI-BamHI fragment containing the 3'-part of the HBsAg gene and the ADH terminator was derived from plasmid pPGT16-3.

Plasmids Pyk3-Pyk6 were prepared as follows.

Plasmid plot5PyksAg51 was digested with BamHI, then with XbaI and the about 580bp fragment containing the Pyk promoter and the 5'-part of the HBsAg gene isolated as above. The 1.1kb BamHI-XbaI fragment, containing the 3'-portion of the HBsAg gene and ADH terminator, was recovered from pPGT16-3, also as above, and the two fragments ligated, digested with BamHI and inserted with different orientations into the BamHI site of pC1/1GAL4/650 (Pyk3, Pyk4). Plasmids Pyk5 and Pyk6 were prepared similarly except that the SalI-XbaI fragment containing the Pyk promoter and 5'-end of the sAg gene was isolated from plot5PyksAg.57 and the XbaI-SalI sAg gene 3'-portion/ADH terminator fusion fragment was derived from plot5sAgtADH and thus both fragments included approximately 6bp of plot5 polylinker sequence. The cassette so formed was then cloned into the SalI site of pC1/1GAL4/370 in opposite orientations.

The six plasmids designated Pyk1-6 (see Fig. 2) are distinguished by Pyk1 having the promoter region proximal to the short pBR322 sequence; Pyk2 having the promoter region proximal to the long pBR322 sequence; Pyk3 having the promoter region proximal to the short pBR322 sequence and distal from the GAL4 sequence; while Pyk4 has the promoter region proximal to the GAL4 region, which in turn is proximal to the long pBR322 sequence; Pyk5 has the promoter region proximal to the GAL4 region which is proximal to the short pBR322 sequence; while Pyk6 has a promoter region distal from the GAL4 region and proximal to the long pBR322 sequence.

These plasmids described above were transformed into S. carlsbergensis strain 2150-2-3 (available from Lee Hartwell, University of Washington) under conventional conditions (Hinnen et al., Proc. Natl. Acad. Sci. USA (1978) 75:1929-1933). Cultures of 50-150ml were grown to mid or late log phase in rich media (YEP) under neutral conditions (3% glycerol, 2% lactic acid), and then inducing conditions (+2% galactose), or

repressing conditions (+2% glucose) for the final 1-2 generations. After lysis with glass beads and clarification of the supernatants by centrifugation, HBsAg expression was determined as described above. The results for the 12 plasmids are set forth in the following Table 1.

Table 1: Expression of HBsAg from Gal Regulated Hybrid Promoters

Construction	YEP + Glycerol / Lactic acid µg sAg/mg protein	YEP + Galactose µg sAg/mg protein	YEP + Glucose	Induction (Gal/glycerol lactic acid)
GAP1	0.04	0.09	0.02	2.0
GAP2	1.65	0.8	1.4	0.5
GAP3	0.25	0.30	---	1.2
GAP4	0.10	0.75	---	7.5
GAP5	0.25	2.1	---	8.4
GAP6	1.55	1.4	1.0	0.9
PYK1	0.10	0.30	0.14	3.0
PYK2	1.65	1.4	1.1	0.85
PYK3	0.10	0.15	---	1.5
PYK4	0.10	1.0	0.05	10.0
PYK5	0.03	1.4	0.02	47.0
PYK6	1.7	1.8	0.9	0.9

0164556

Construction of pPGAP.

A yeast expression vector was prepared calle pPGAP having a polyrestriction site linker between the GAPDH terminator and short promoter region. Plasmid plotSpGAP128 was digested with BamHI and TaqI to yield an approximately 390bp BamHI-TaqI fragment having the -400 to -26bp of the GAPDH promoter. The BamHI-TaqI fragment was ligated to a synthetic fragment having th following sequence:

$$CGA_2TA_3(CA)_3TA_3CA_3CACCATG_3A_2T_2CGT_2AG_2$$
$$T_2AT_3(GT)_3AT_3GT_3GTGGTAC_3T_2A_2GCA_2TC_2AGCT$$

to provide a BamHI-SalI fragment, which was digested with BamHI and SalI and used to replace the BamHI-SalI fragment of BamHI-SalI digested pBR322 treated with alkaline phosphatase. After ligation, the plasmid pGAPNRS was obtained which was digested with BamHI and SalI to provide a 400bp BamHI-SalI fragment which was gel isolated. This fragment was ligated to an about 900bp SalI-BamHI fragment containing the GAPDH termina- tor region and a short segment of 3' coding region and the resulting 1.4kb BamHI-BamHI fragment digested with BamHI. The SalI-BamHI GAPDH terminator fragment was obtained by SalI and BamHI digestion of pGAP2, a plasm prepared by insertion of an about 3.3kb BamHI fragment containing the GAPDH gene GAP49 (Holland and Holland, supra) into the BamHI site of pBR322. Plasmids pGAP2 and pGAP1 were obtained as follows: A yeast gene li- brary was prepared by inserting fragments obtained aft partial digestion of total yeast DNA with restriction endonuclease Sau3A in lambda-phage Charon 28 (Blattner et al., Science (1977) 196:161-169). The phage librar was screened with DNA complementary to the yeast GAPDH mRNA and the yeast GAPDH gene from one of these clones was subcloned as either an about 3.3kb BamHI fragment in the BamHI site of pBR322 (pGAP-2) or as an about

2.1kb HindIII fragment in the HindIII site of pBR322 (pGAP-1).

pBR322 was digested with EcoRI and SalI, the termini blunt-ended and ligated to BamHI linkers, followed by BamHI digestion and the BamHI-BamHI 3.8kb fragment gel isolated, recircularized by self-ligation, cloned and designated pBRΔR1-Sal. The 1.4kb BamHI-BamHI fragment was inserted into the BamHI-digested, alkaline phosphatase treated pBRΔR1-Sal vector to provide the plasmid pPGAP of about 5.3kb with the orientation in the opposite direction of the amp^r.

The plasmid phSOD was prepared as follows:

Molecular cloning of hSOD cDNA.

Total RNA was prepared from an adult human liver by the guanidinium thiocyanate/lithium chloride method (Cathala et al., DNA (1983) 2:329-335). polyA RNA was used to synthesize double-stranded cDNA (Maniatis et al., Molecular Cloning, 213-242, Cold Spring Harbor, 1982) and this was passed over a Sepharose CL4B column to enrich for cDNAs of greater than 350bp (Fiddes and Goodman, Nature (1979) 281:351-356). The cDNA was inserted at the PstI site of plot4, a pBR322 derivative having the following sequence replacing the PstI-EcoRI site.

```
          PstI     HinfI                    AluI
     1    GGTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
          ACGTCCACTTAGGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAG
               HphI

                                                    HindIIAluI
    60    ACAATTCCACACATTATACGAGCCGATGATTAATTGTCAACAGCTCATTTCAGAATATTT
          TGTTAAGGTGTGTAATATGCTCGGCTACTAATTAACAGTTGTCGAGTAAAGTCTTATAAA

                                    EcoRI
   120    GCCAGAACCGTTATGATGCGG
          CGGTCTTGGCAATACTACGCCTTAA
```

The cDNA insertion employed the oligo-dG:dC tailing method (Maniatis et al., supra). E. coli strain D1210 was transformed with this mixture and transformants selected on L-agar containing 10µg/ml tetracycline (Kushner, S.R. (1978) In: Genetic Engineering, eds. Boyer, H.B. and Nicosia, S., (Elsevier/North Holland, Amsterdam) p. 17). Plasmid DNA constituting a liver cDNA library was prepared (Maniatis et al., Molecular Cloning, pp. 86-94, Cold Spring Harbor 1982) directly from approximately 62,000 recombinant colonies plated at a density of approximately 3,000 colonies per 9cm diameter Petri dish.

Isolation of r-hSOD clones.

Strain D1210 was retransformed with the liver cDNA library and about 40,000 clones were grown on nine 14cm diameter Petri dishes. After transfer of the colonies to nitrocellulose paper and chloramphenicol amplification of plasmid DNA, the cells were lysed and the filters prepared for hybridization (Ish-Horowicz and Burke, Nucleic Acids Research (1981) 9:2989-2998). Oligonucleotide probes were employed for screening by hybridization, with the probes consisting of enzymatically-radiolabeled, chemically-synthesized DNA molecules complementary to the mRNA encoding amino acid residues 19 to 24 of the protein (Jabusch et al., supra.; Barra et al., supra.); the mixture had the following sequences:

```
3' TTA AAA CTT GTT TTT CT 5'
     G   G   C   C   C
```

where all of the indicated possibilities for encoding the peptide sequence were prepared (32-fold degenerate).

The probes were labeled with ^{32}P to a specific activity of $1-3\times10^{8}$ cpm/µg and Millipore (0.45µm) filtered before use. Filters were prehybridized for 6hrs at 30°C in 4x SSC, 2x Denhardts's solution, 40mM sodium

phosphate, pH 7.5, 300µg/ml sonicated salmon testes DNA. Hybridization was for 20hrs at 30°C in the same solution containing $2x10^6$cpm/ml hSOD DNA probe (residues 19-24). Filters were washed in 4x SSC, once for 15min at r.t. and twice for 15min at 30°C, blotted dry and autoradiographed with an intensifying screen for 24hrs at -70°C.

Areas on the master plates that corresponded to duplicate positive signals were picked into L-broth and plasmid DNA prepared by the miniscreen procedure (Maniatis et al., Molecular Cloning, 178, 368-369, Cold Spring Harbor 1982). This DNA was cut with PstI and subjected to Southern blot analysis (Southern, J. Mol. Biol. (1975) 98:503-517) hybridizing initially with the previous labeled probes (amino acid residues 19-24) and then with additional radiolabeled probes derived from amino acid residues 109-114 and having the following sequences (all possible variations, 72-fold degenerate) present as a mixture:

```
3' CTA GTA ACA TAA TAA CC 5'
     G   G   G   G   G
                 T   T
```

One plasmid pool (pSOD1) contained a cDNA insert of 520bp that hybridized with both probes and after colony purification, plasmid DNA was prepared from this clone and sequenced by the method of Maxam and Gilbert (Proc. Natl. Acad. Sci. USA (1977) 74:560-564). The hSOD cDNA clone pSOD1 constitutes the coding region for amino acids 10-153 of hSOD, a single translational stop codon and a 3' untranslated region. Therefore, in the expression vector construct, the base sequence of the region encoding amino acids 1-9 is derived from the published amino acid sequence of hSOD (Jabusch et al., supra; Barra et al., supra) and synthesized chemically as a

part of the variable linker segment (see discussion relating to Fig. 3).

Construction of plot5 derivatives containing r-hSOD.

The synthetic DNA molecules F(26), C(16), B(31), D(11), E(13) and 4(24) shown in Fig. 3, were synthesized by the phosphoramidite method.

The single strand 4(24) was prepared by using all four bases, at each site where X is indicated. Furthermore, silica was withdrawn from the synthesis of the 24mer, such that single-stranded 21mers, 22mers, and 23mers are obtained in addition to the 24mers. After removal from the silica support, the four mixtures are combined in appropriate proportions to provide for equimolar amounts of each of the possible single strands. This mixture was treated as a single product in the subsequent steps.

Molecules F(26), C(16), B(31) and D(11) were mixed together in equimolar amounts and 10µg phosphorylated using T4 polynucleotide kinase. After phenol-ether extraction, the additional non-phosphorylated synthetic DNA molecules 4(24) and E(13) were added, such that all fragments were equimolar. The equimolar mixture contained 13µg of DNA in 133µl of 0.3x kinase buffer.

After annealing by cooling at a uniform rate from 70°C to 20°C over 60min, the single strands were ligated together with T4 ligase in 200µl ligation mix at 14°C for 4hrs, phenol-chloroform extracted, ethanol precipitated and the 5'-ends of 4(24) and E(13) phosphorylated using T4 polynucleotide kinase (Maniatis et al., supra). Preparative polyacrylamide gel electrophoresis was used to isolate the completely ligated 53bp material having 5'- and 3'-overhangs.

The above purified fragment mixture was then ligated to the 460bp TaqI-PstI segment of the hSOD cDNA as shown in Fig. 3. This segment was itself constructed by isolating the 454bp TaqI-AluI hSOD fragment, making

it flush-ended using Klenow and inserting it into plot5 between its EcoRI and SalI sites which had been similarly made flush-ended. After preparation of plasmid DNA from this recombinant, the 460bp TaqI-PstI hSOD fragment was isolated by preparative polyacrylamide gel electrophoresis. After extraction and precipitation, the 515bp fragment resulting from the joining of the synthetic fragment to the 460bp TaqI-PstI hSOD fragment was blunt-ended (525-528bp) and then digested with SalI and the resulting 519-522bp hSOD fragment isolated by polyacrylamide gel electrophoresis. This fragment was then inserted into plot5 which had been digested with PvuII and SalI and then treated with alkaline phosphatase. The resulting plasmids were used to transform strain D1210. Recombinants obtained after transformation of strain D1210 were selected on L-agar containing 100µg/ml ampicillin to give a set of clones, which were screened for an NcoI site. One was selected and designated phSOD.

Construction of a yeast vector for SOD expression.

The plasmid phSOD was digested with NcoI and SalI and a 550bp fragment obtained, which included 1 nucleotide untranslated at the 5'-terminus and the entire coding region for hSOD. pPGAP was digested with NcoI and SalI followed by treatment with alkaline phosphatase and the SalI-NcoI fragment substituted for the NcoI-SalI fragment in pPGAP to provide pPGAPSOD. BamHI digestion of pPGAPSOD resulted in a 2kb fragment which was gel isolated and inserted into the BamHI site of pC1/1 and pC1/1 GAL4/370. These plasmids were transformed into yeast strain 2150-2-3 as described previously, with the results of expression set forth in the following Table 2.

Table 2: Expression of Human SOD in Yeast Strain 2150

Plasmid	Carbon Source	SOD[2] µg/mg protein
pC1/1	g, L[1]	0
pC1/1GAPSOD	g, L	148
pC1/1GALGAPSOD	g, L	0.4
	gal	68

[1] All cultures grown in Minus Leucine media with 2% lactic acid, 3% glycerol with or without 2% galactose to late log or early stationary phase.

[2] Determined by RIA.

hSOD levels were measured using a standard radioimmunoassay with iodinated authentic hSOD as standard. Constitutive synthesis from the GAP promoter leads to very high levels of hSOD production, of the order of 10-30% of the total cell protein. The induction with galactose works almost as well, yielding about 7% of the cell protein as hSOD.

Construction of pYASI1

This yeast expression plasmid contains the hSOD gene fused to the amino terminus of the human proinsulin gene, with a methionine codon at the junction between both genes. The fusion gene is under control of the hybrid inducible ADH2-GAP (yeast alcohol dehydrogenase 2) promoter and the GAP terminator. An about 3kbp BamHI expression cassette was constructed.

The ADH2 portion of the promoter was constructed by cutting a plasmid containing the wild type ADH2 gene (plasmid pADR2, see Beier and Young, Nature (1982) 300:724-728), a pBR322 derivative, which contains the yeast centromere gene from chromosome 3 (CEN3) (U.S. Patent No. 4,464,472), the Trp1 gene and the yeast ADHII

gene with the restriction enzyme EcoR5, which cuts at a position +66 relative to the ATG start codon, as well as in two other sites in pADR2, outside of the ADH2 region in the CEN3 and Trp1 gene. The resulting mixture of a vector fragment and two smaller fragments was resected with Bal31 exonuclease to remove about 300bp. Synthetic XhoI linkers were ligated onto the Bal31 treated DNA. The resulting DNA linker vector fragment was separated from the linkers by column chromatography, cut with the restriction enzyme XhoI, religated and used to transform E. coli to ampicillin resistance. The positions of the XhoI linker additions were determined by DNA sequencing. One plasmid which contained an XhoI linker located within the 5' non-transcribed region of the ADH2 gene (position -232 from ATG) was cut with the restriction enzyme XhoI, treated with nuclease S1, and subsequently treated with the restriction enzyme EcoRI to create a linear vector molecule having one blunt end at the site of the XhoI linker and an EcoRI end at the pBR322 proximal terminus.

The GAP portion of the promoter was constructed by cutting plasmid pPGAP (supra) with the enzymes BamHI and EcoRI, followed by the isolation of the 0.4Kbp DNA fragment. The purified fragment was cut with the enzyme AluI to create a blunt end near the BamHI site.

Plasmid pJS014 was constructed by the ligation of the AluI-EcoRI GAP promoter fragment to the ADH2 fragment present on the linear vector described above.

Plasmid pJS104 was digested with BamHI (which cuts upstream of the ADH2 region) and with NcoI (which cuts downstream of the GAP region). The about 1.3Kbp fragment containing the ADH2-GAP promoter was gel purified and ligated to an about 1.7Kbp fragment containing the hSOD-proinsulin fusion DNA sequences and GAP terminator present in pYSI1. This 3Kbp expression cassette was cloned into BamHI digested and phosphatase treated pC1/1 to yield pYASI1.

Plasmid pYSI1 was constructed as follows. Three fragments were employed which involve a 454bp NcoI-Sau3A isolated from phSOD, where the fragment includes the entire coding sequence for human superoxide dismutase (hSOD) with the exception of the last three 3'- codons; a 51bp Sau3A-HindIII synthetic adapter, which codes for the last three codons of hSOD, methionine, and the first 14 codons of proinsulin; and a 231bp HindIII-SalI fragment, isolated from pINS5 (a pBR322-derived vector which contains a proinsulin coding sequence chemically synthesized according to the amino acid sequence reported by Bell et al., Nature (1979) 282:525-527), which encodes proinsulin excepting the first 14 amino acids. These fragments were ligated together and introduced into the plasmid pPGAP, which had been previously digested with NcoI and SalI and alkaline phosphatase treated. The resulting plasmid pSI1 was digested with BamHI to provide an expression cassette which was cloned into plasmid pC1/1 to yield pYSI1.

Cloning of alpha-1-antitrypsin.

A cDNA library was made from 10µg of polyA$^+$ RNA isolated from a part of a human liver. This library was prepared by oligo-dT priming of the first cDNA strand and self-priming of the second cDNA strand. The ds cDNA was size fractionated on a Sepharose CL4B column and those molecules greater than 300bp isolated. This fraction was treated with nuclease S1 and tailed with dCTP, using terminal transferase. The tailed cDNA was annealed to pBR322 which had been digested with PstI and tailed with dGTP. Transformation of E. coli HB101 yielded 60,000 colonies, where greater than 90% of the clones were recombinant.

Two synthetic oligonucleotide probes were used to isolate the alpha-1-antitrypsin (α_1-AT) cDNA, the first probe corresponding to amino acid residues

344-350 near the C-terminus of the protein was used to probe 5,000 colonies and the second probe, corresponding to amino acid residues -23 to -17 (+1 being the first nucleotide of the first codon of the mature α_1-AT) of the signal peptide, was used to probe 25,000 colonies. The probe sequences were taken from the partial nucleotide sequence described by Kurachi et al., Proc. Natl. Acad. Sci. USA (1981) 78:6826; Leicht et al., Nature (1982) 297:655). Approximately 3% of the colonies hybridized to the C-terminal probe and four hybridized to the N-terminal probe. The four N-terminal clones and 12 C-terminal clones were isolated and subjected to restriction analysis. From these, three overlapping clones which cover the entire cDNA were subjected to further study and were used to construct the full length cDNA clone.

The entire sequence of a composite full length cDNA derived from the three plasmids is as follows:

```
                                                
```

LEGEND

Nucleotide and predicted amino acid sequences of α_1-AT cDNA. The reactive center met-ser at positions 358-359 is boxed. Subscripts to amino acids in parentheses identify differences between the subject protein sequence and those derived from (a) protein sequencing (Carrell et al., 1982), (b) the cDNA of Woo et al., [see Carrell et al., 1982]), and (c) the cDNA of Bollen et al., 1983. The synthetic DNA molecules used in the construction of the BamHI to SalI fragment encoding the mature protein are shown as are the cDNA restriction sites used in this construction.

The above sequence was determined using the dideoxy sequencing method of Sanger et al., Proc. Natl. Acad. Sci. USA (1977) 74:5463, in the M13 vectors of Messing et al., Nucleic Acids Res. (1981) 9:309. The differences at the nucleotide and amino acid level from the published cDNA sequences are shown.

Construction of the full length clone for expression of yeast began with three fragments isolated from cDNA clones: 1) a 630bp BamHI-BstEII fragment; 2) a 450bp BstEII-AvaI fragment; and 3) an 85bp AvaI-HinfI fragment. A synthetic adapter was employed having the following sequence:

$$A_2TC_3AC_3A_5TAG$$
$$G_3TG_3T_5ATCAGCT$$

Approximately two pmoles of fragments 1 and 2 were ligated together and after removal of the ligase, digested with BamHI and AvaI. Fragment 3 and the synthetic adapter were ligated and digested with AvaI and SalI and the two resulting fragment mixtures were ligated followed by digestion with BamHI and SalI. Fragments migrating upon electrophoresis in the region of about 1000-1400bp were isolated and cloned by substitution into BamHI and

SalI digested and alkaline phosphatase treated pBR322. The resulting plasmid is referred to as pATi.

Plasmid pPGAP was digested with NcoI, followed by blunt-ending, followed by SalI digestion and treatment with alkaline phosphatase. The NcoI-SalI fragment was substituted with an approximately 1250bp blunt-ended (BamHI)-SalI fragment obtained from plasmid pATi, by BamHI digestion, blunt ending, and SalI digestion. This was inserted into the pPGAP vector to produce the plasmid pGAPATi, a 6.6kb plasmid, which was digested with NcoI and BamHI and a 2.3kb NcoI-BamHI fragment obtained having the α_1-AT gene and the GAPDH terminator and approximately 400bp BamHI-NcoI fragment obtained having the GAPDH promoter. These fragments were ligated together and inserted into the BamHI site of pC1/1. The plasmids pC1/1GAPATi8 and pC1/1GAPATi9 were obtained with the orientation of expression clockwise in the former and counterclockwise in the latter, with amp^r being in the counterclockwise direction. These plasmids were transformed in S. cerevisiae AB103 (A.T.C.C. No. 20658, deposited January 5, 1983) by standard methods, selecting for leucine prototrophy and grown as described above. Yeast extracts were prepared by lysis with glass beads and the α_1-AT activity determined by inhibition of human leukocyte elastase.

Assays contained in 1ml:0.1-0.2 human leukocyte elastase (HLE); 0.1mM MeO-Suc-Ala-Ala-Pro-Val-p-nitroanilide (Beatty et al., J. Biol. Chem. (1980) 255:3931); 50mM Tris, pH 8, 0.5M NaCl, and the indicated amounts of yeast extract or human α_1-AT. Assays were initiated by the addition of elastase, incubated at 28°C for 15min, terminated by the addition of 100µl of 8N acetic acid and the absorbance at 410nm determined. Typical results are shown in the following Table 3.

Table 3

Plasmid	Strain	Amt.Extract (µl)	Amt.HLE (µg)	Amt.Protein (µg)	% Elastase Activity	% α_1-AT*
pC1/1GAPATi8	AB103	5.0	0.1	50.0	40	0.17
		10.0	0.1	100.0	26	0.11
pC1/1GAPATi9	AB103	0.25	0.1	2.3	89	0.7
		1.0	0.1	9.1	26	1.2
pC1/1GAPATi9	AB110	0.2	0.2	2.9	39	6.1
		0.4	0.2	5.8	14	4.3

* Calculation based upon the Mol. wt. of HLE (29kD), the amount of protein added and the degree of inhibition.

0164556

The above data demonstrate that plasmids having the orientation of the expression cassette in the counter-clockwise orientation, the promoter proximal to the long sequence of pBR322, make 10-20 times more α_1-AT than the same cassette in the other orientation.

Yeast strain AB110.

Yeast strain 2150-2-3 was crossed with a yeast strain AB103 transformant containing pC1/1GAPATi9. The diploids were sporulated and the tetrads dissected. Strains were maintained on leucine selective plates in order to ensure maintenance of the plasmid, since the parents are auxotrophs. A series of colonies were screened for their genotype with respect to a number of markers. The most vigorous strains were selected and cultures grown on leucine selective media. The best strain was designated AB110 (pC1/1GAPATi9), gave 6-7.5% of the total cell protein as α_1-AT as shown in the above Table 3. The strain AB110 has the following genotype: Matα, ura3-52, leu2-04 or both leu2-3 and leu2-112, pep4-3, his4-580 (cir°).

Phosphate induction.

Plasmid pPGT80 was digested with BamHI, the ends blunt-ended, followed by digestion with XbaI and the 500bp fragment containing the GAPDH promoter and 5'-end of the sAg gene isolated.

The PHO5 gene was isolated from a yeast genomic library employing an oligonucleotide probe 5'-GGCACTCACACGTGGGACTAG-3' derived from the published partial sequence (Meyhack et al., The EMBO Journal (1982) 1:675-680). A subfragment of this clone containing 550bp of the 5'-untranslated region and approximately 80bp of coding sequence was subcloned as a BamHI-SalI substitution in pBR322 to provide pPHO5. This fragment has the sequence 5'-ATGTTTAAA-3', encoding the first three amino acids, the second and third codons

specifying an AhaIII site. The plasmid pHBS6 was digested with NcoI, blunt-ended, followed by digestion with BamHI and treatment with alkaline phosphatase. The PHO5 promoter region was obtained by digesting the pPHO5 plasmid with AhaIII, resecting the resulting fragment with Bal31 for a short time, followed by digestion with BamHI and isolation of a 500-550bp BamHI blunt-ended fragment. This fragment was employed for substitution of the NcoI-BamHI fragment from pHBS6 and was screened for regeneration of the NcoI restriction site to provide plasmid pHBS6PHO5/1.

Plasmid pHBS6PHO5/1 was digested with BstEII which cleaves at position -175 in the PHO5 promoter. This molecule was blunt-ended, digested with SalI and the 650bp fragment having the 5'-portion of the promoter domain, containing 275bp of pBR322 and 375bp of the PHO5 promoter region isolated. This fragment was ligated with the blunt-ended (BamHI)-XbaI fragment obtained from digestion of pPGT80 with BamHI, blunt ending, followed by XbaI digestion. After digesting the ligated fragment with SalI and XbaI, the resulting fragment was then substituted into pPGT16-3 which had been digested with SalI and XbaI and treated with alkaline phosphatase. The resulting plasmid pPHO5PGT80 had a cassette comprising the PHO5 regulatory region, the GAPDH promoter, the sAg gene and the ADH terminator. This cassette was excised from the plasmid by BamHI digestion, whereby a 1.8kb BamHI-BamHI fragment was gel isolated and ligated into the BamHI site of BamHI digested and alkaline phosphatase treated pC1/1 to provide plasmids PHO5GAP1 and PHO5GAP2 where the PHO5 was distal and proximal to the long pBR322 sequence, respectively.

The two plasmids were transformed into yeast strain 2150-2-3 as described above and grown in rich media as described above for 8 to 10 generations in either high (7mM) or low (0.2mM) phosphate. Samples were harvested in late log phase and HBsAg determined

as described previously. The results are shown below in Table 4.

Table 4: Regulation of HBsAg Production in Yeast using a Hybrid PHO5/GAPDH Promoter.

Construction	High Phosphate (7mM) (sAg µg/mg protein)	Low Phosphate (0.2mM)	Induction low / high
PHO5GAP-1	0.08	0.95	12.0
PHO5GAP-2	0.27	0.40	1.5

From the above results, it is evident that effective regulation with phosphate is obtained, with one orientation being superior to the other.

It is evident from the above results, that highly efficient expression can be obtained, either constitutive or regulated, by providing for truncated promoter regions of yeast glycolytic enzyme gene promoters, employing the 3' domain proximal to the coding region of the gene in conjunction with a 5'-portion or second domain of the promoter region of a yeast gene subject to inducible regulation by a nutrient, e.g., carbon source or phosphate, temperature, or other externally controllable source or condition. Alternatively, the second domain may be replaced by prokaryotic sequences of at least about 1kb or greater, which provide for constitutive enhancement in the absence of the second domain of the promoter region. Thus, a wide variety of genes exogenous to yeast may be expressed in high yield in high percentages of the total protein of the yeast host.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. In a method for obtaining expression of a foreign structural gene in a yeast host employing hybrid constructions involving transcriptional and translational regulatory signals recognized by said host joined to a foreign gene for transcriptional and translational regulation of expression in said host, wherein said construction is introduced into said host and said foreign gene is expressed to produce the encoded polypeptide,

the improvement which comprises:

employing for transcriptional regulation, a DNA sequence having first and second domains,

said first domain being proximal to said structural gene, being of less than about 600bp and having a sequence which includes the RNA polymerase binding site and transcription initiation site of a yeast glycolytic enzyme gene; and

said second domain distal to said structural gene providing for enhanced efficiency of transcription and being the regulatory region of a regulated yeast gene other than the natural gene of said first domain or a prokaryotic sequence of at least about 500bp.

2. A method according to Claim 1, wherein said yeast glycolytic enzyme gene of said first domain is glyceraldehyde-3-phosphate dehydrogenase.

3. A method according to Claim 2, wherein said regulatory region is the GAL1 or GAL10 regulatory region.

4. A method according to Claim 2, wherein said regulatory region is the PHO5 regulatory region.

5. A method according to Claim 1, wherein said yeast glycolytic enzyme of said first domain is pyruvate kinase.

6. A method according to any of Claims 1, 2, 3, 4, or 5, wherein said polypeptide is Hepatitis B surface antigen, superoxide dismutase, or α_1-antitrypsin.

7. In a method for obtaining expression of a foreign structural gene in a yeast host employing a hybrid construction involving transcriptional and translational regulatory signals recognized by said host joined to a foreign gene for transcriptional and translational regulation of expression in said host, wherein said construction is introduced into said host and said foreign gene is expressed to produce the encoded polypeptide, the improvement which comprises employing for transcriptional regulation, a DNA sequence having first and second domains,

said first domain being proximal to said structural gene, being of about 200-500bp, being derived from the 5'-untranslated region of the yeast glyceraldehyde-3-phosphate dehydrogenase or pyruvate kinase gene and extending downstream to at least nucleotide -10; and

said second domain distal to said structural gene, being a prokaryotic sequence of at least about 1000bp, wherein said yeast cells containing said construct are grown to high density; and

isolating said polypeptide.

8. In a method for obtaining expression of a foreign structural gene in a yeast host employing hybrid constructions involving transcriptional and translational regulatory signals recognized by said host joined to a foreign gene for transcriptional and translational regulation of expression in said host, wherein

said construction is introduced into said host and said foreign gene is expressed to produce the encoded polypeptide, the improvement which comprises employing for transcriptional regulation, a DNA sequence having first and second domains,

said first domain being proximal to said structural gene, being of about 200-500bp and including the RNA polymerase binding site and transcription initiation site of the yeast glyceraldehyde-3-phosphate dehydrogenase or pyruvate kinase gene; and

said second domain distal to said structural gene having a sequence providing for regulation of the acid phosphatase PHO5 gene or the GAL1 or GAL10 gene,

growing said host containing said construct in a nutrient medium to high density of said host under non-inducing conditions;

changing the conditions to inducing conditions to induce expression of said structural gene; and

isolating the polypeptide.

9. A DNA construct comprising in the downstream direction a yeast promoter region having first and second domains, an insertion sequence of from about 4-100bp having at least one unique restriction site for the construct, and a termination region, wherein said two domains are a distal domain of at least about 100bp of a DNA sequence providing for regulation of a yeast gene or a prokaryotic sequence of at least about 500bp and a proximal domain which includes the RNA polymerase binding site and the transcription initiation site of a yeast glycolytic enzyme gene.

10. A DNA construct for expression of a structural gene foreign to yeast, comprising in the downstream direction:

a promoter region comprising two domains and providing for transcriptional and translational signals for initiation in yeast;

a structural gene under regulatory control of said promoter region; and

a termination region providing for transcriptional and translational termination in yeast;

wherein the distal domain comprises a sequence of at least about 100bp including a sequence for regulation of a regulatable yeast gene or a prokaryotic sequence of at least about 500bp and the proximal domain includes an RNA polymerase binding site and transcription initiation site of a yeast glycolytic enzyme gene.

11. A construct according to Claim 10, wherein said gene is α_1-antitrypsin, superoxidase dismutase or Hepatitis B surface antigen.

12. A modified yeast host containing as an extrachromosomal element a plasmid vector having a replication system, at least one marker, and a cassette comprising in the downstream direction of transcription, a promoter region functional in yeast for transcriptional and translational initiation, a structural gene foreign to said yeast under the transcriptional and translational control of said promoter region, and a terminator region functional in said yeast, wherein said promoter region has two domains, a distal domain distant from said gene of at least about 100bp comprising a sequence providing for regulation of a regulated yeast gene or a prokaryotic sequence of at least about 500bp, and a proximal domain proximal to said gene, comprising the RNA polymerase binding site and transcription initiation site of a yeast glycolytic enzyme gene and of about 200 to 600bp and extending to at least the -20nt of said yeast glycolytic enzyme gene.

13. A yeast cell according to Claim 12, wherein the yeast glycolytic enzyme gene for said proximal domain is glyceraldehyde-3-phosphate dehydrogenase or pyruvate kinase.

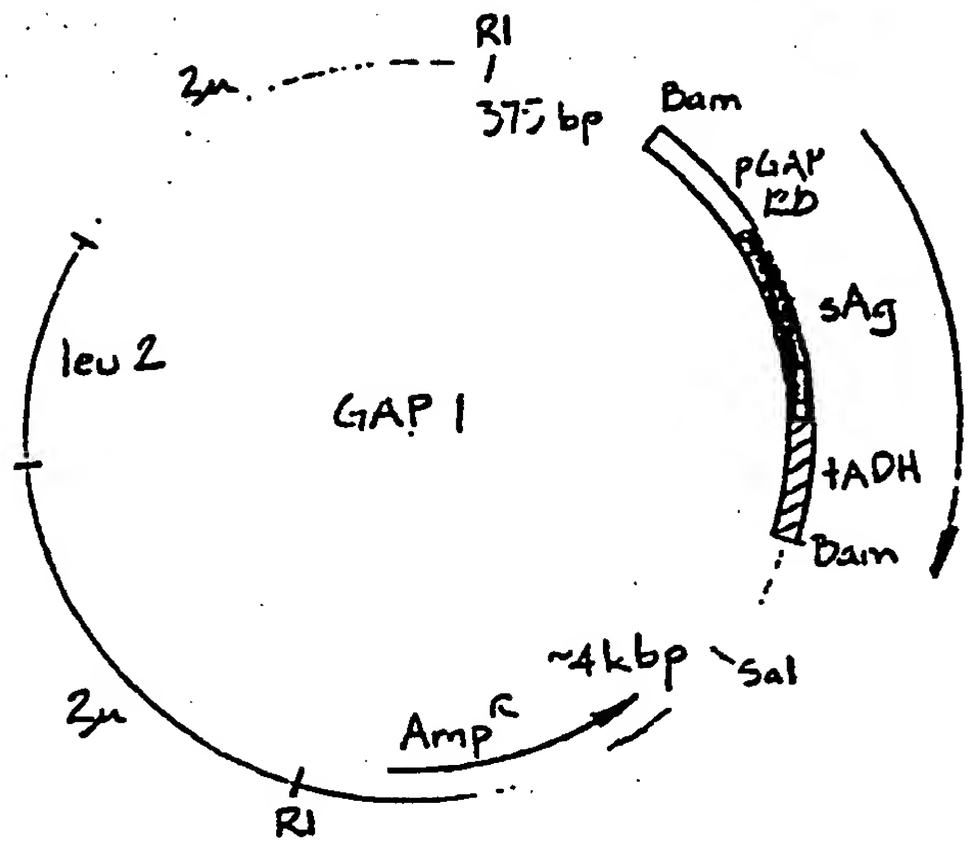
14. A method for producing a foreign protein which comprises:

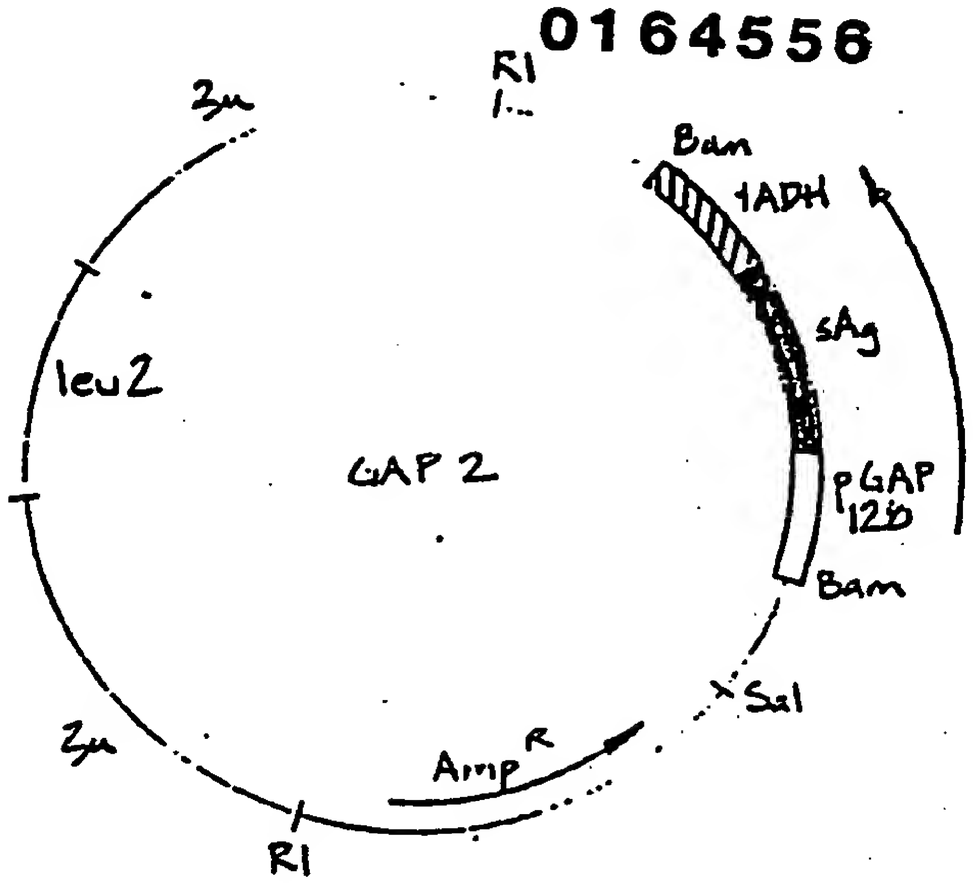
growing cells according to Claim 12, wherein said foreign protein is produced.

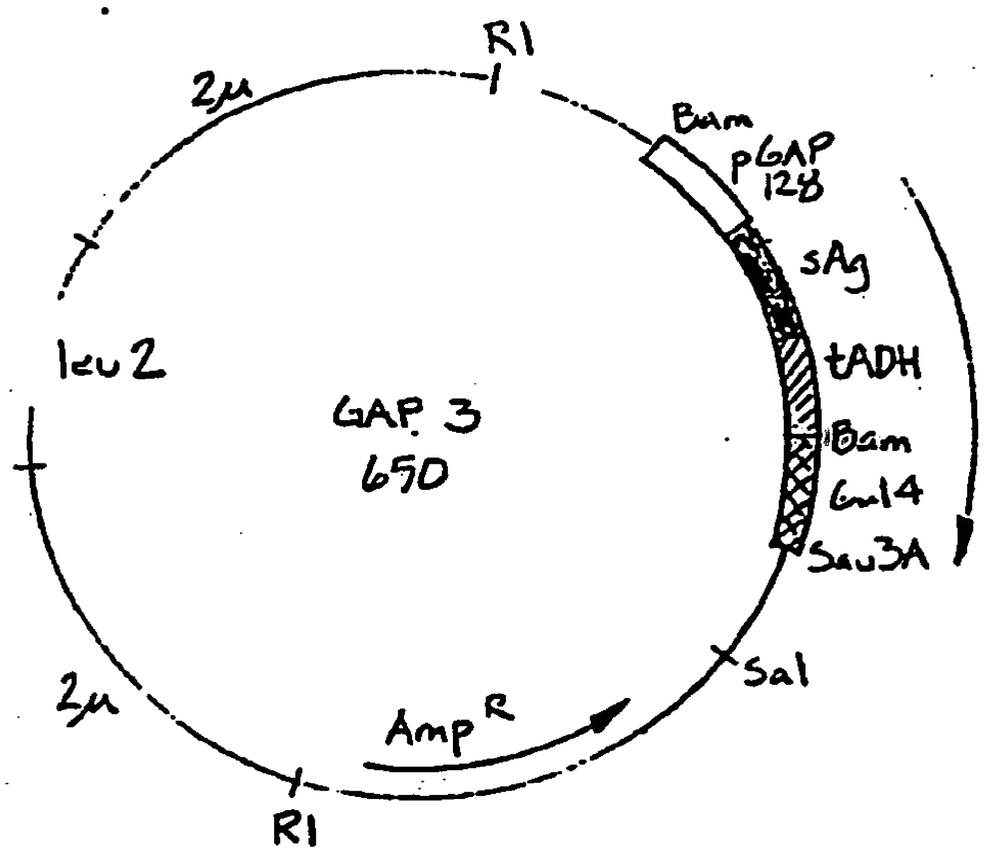
15. A method according to Claim 14, wherein said foreign protein is α_1-antitrypsin, human superoxide dismutase or Hepatitis B surface antigen.

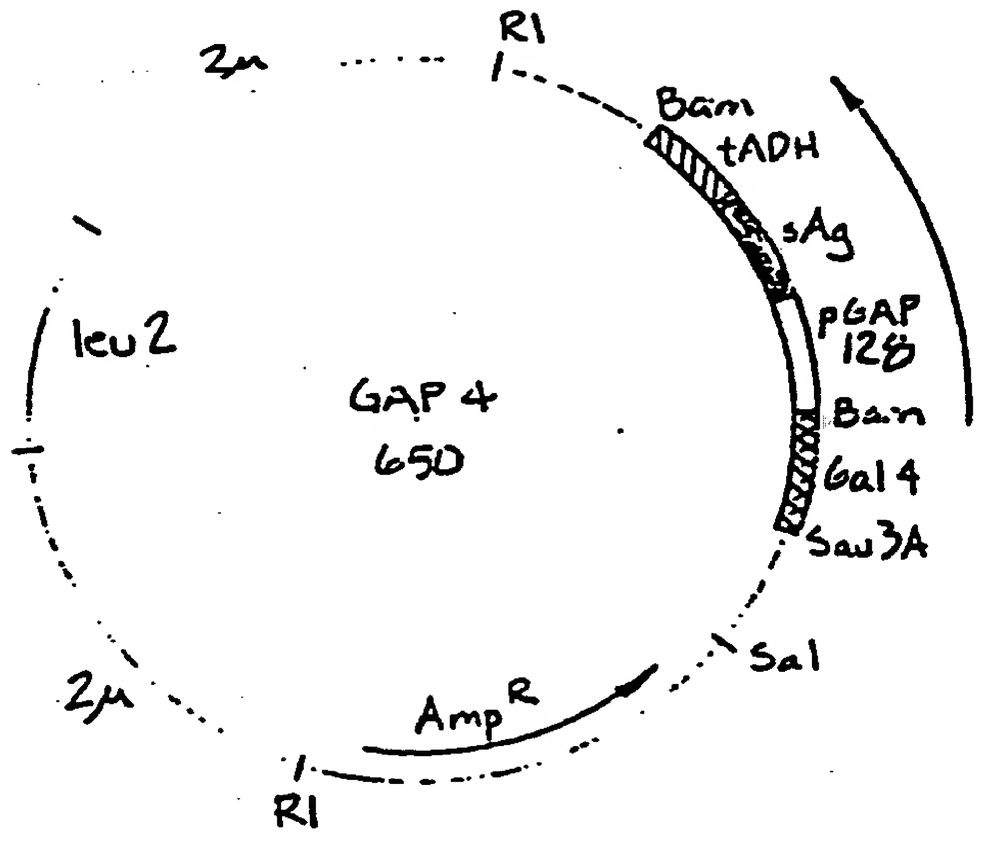
16. A method according to Claim 2, wherein said regulatory region is the ADHII regulatory region.

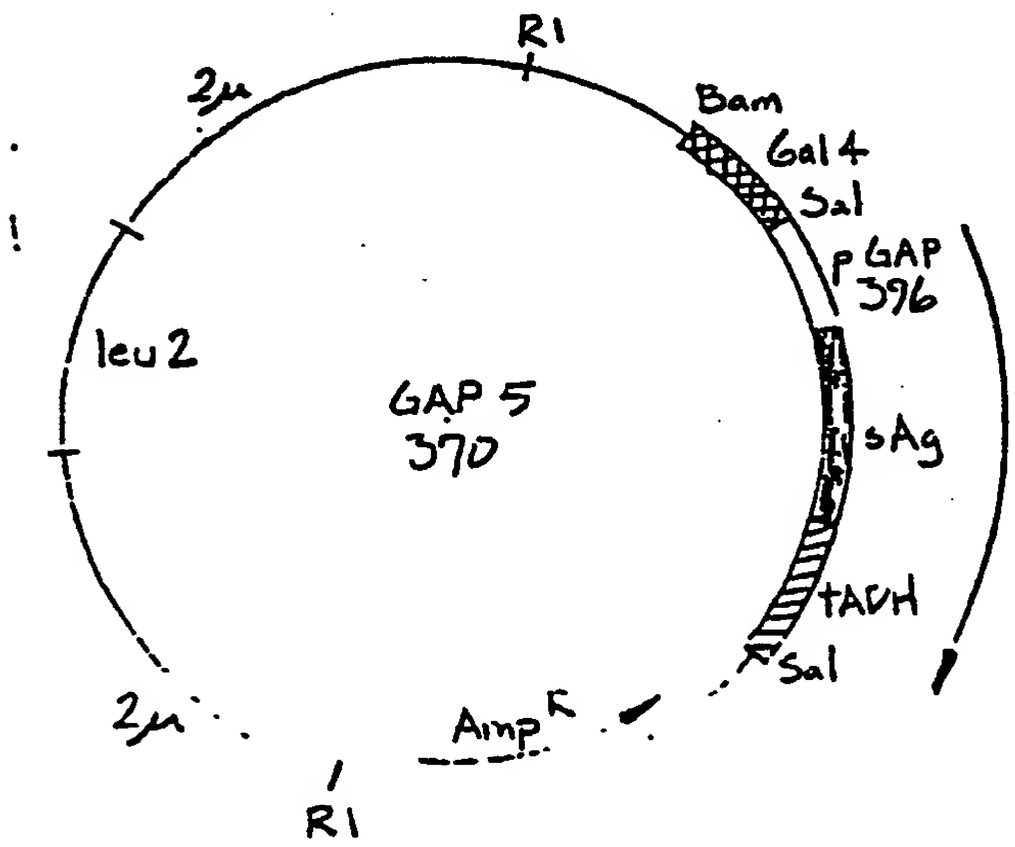
0164556











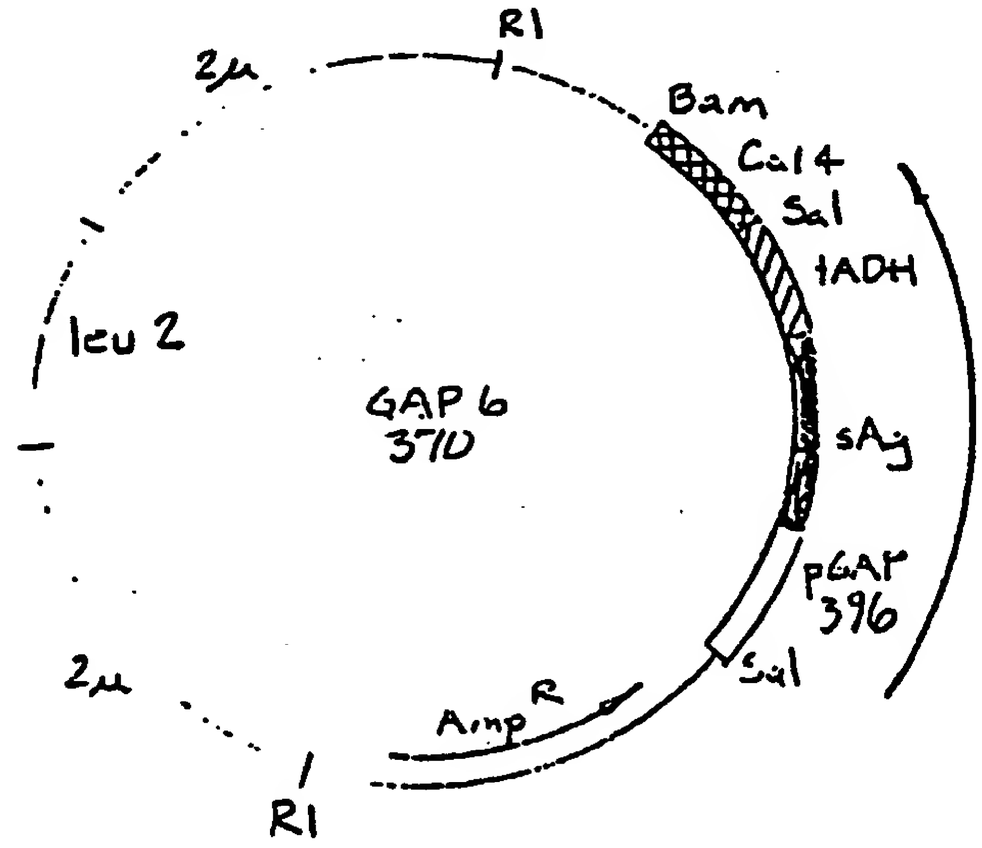


FIG. 1

0164556

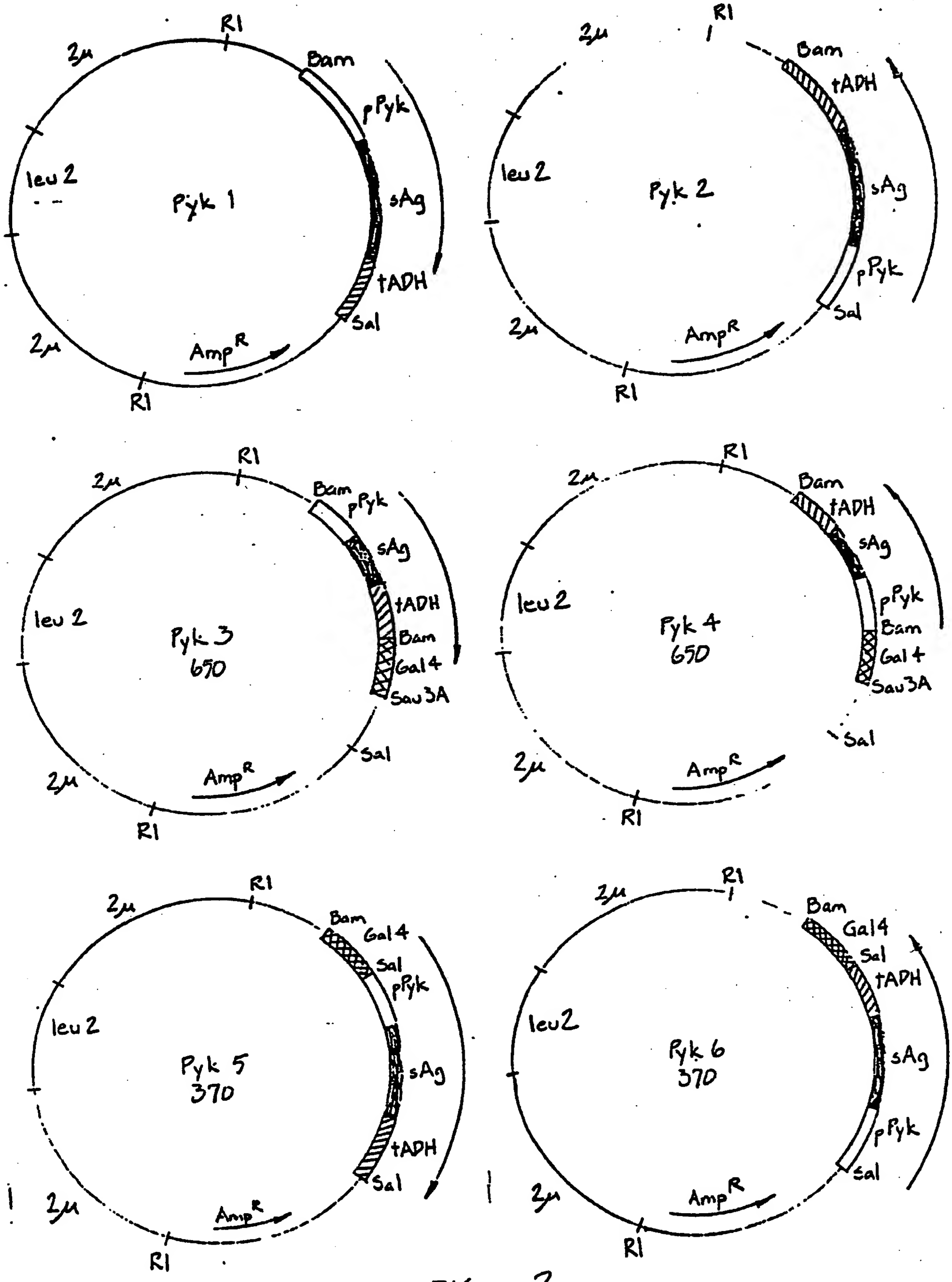
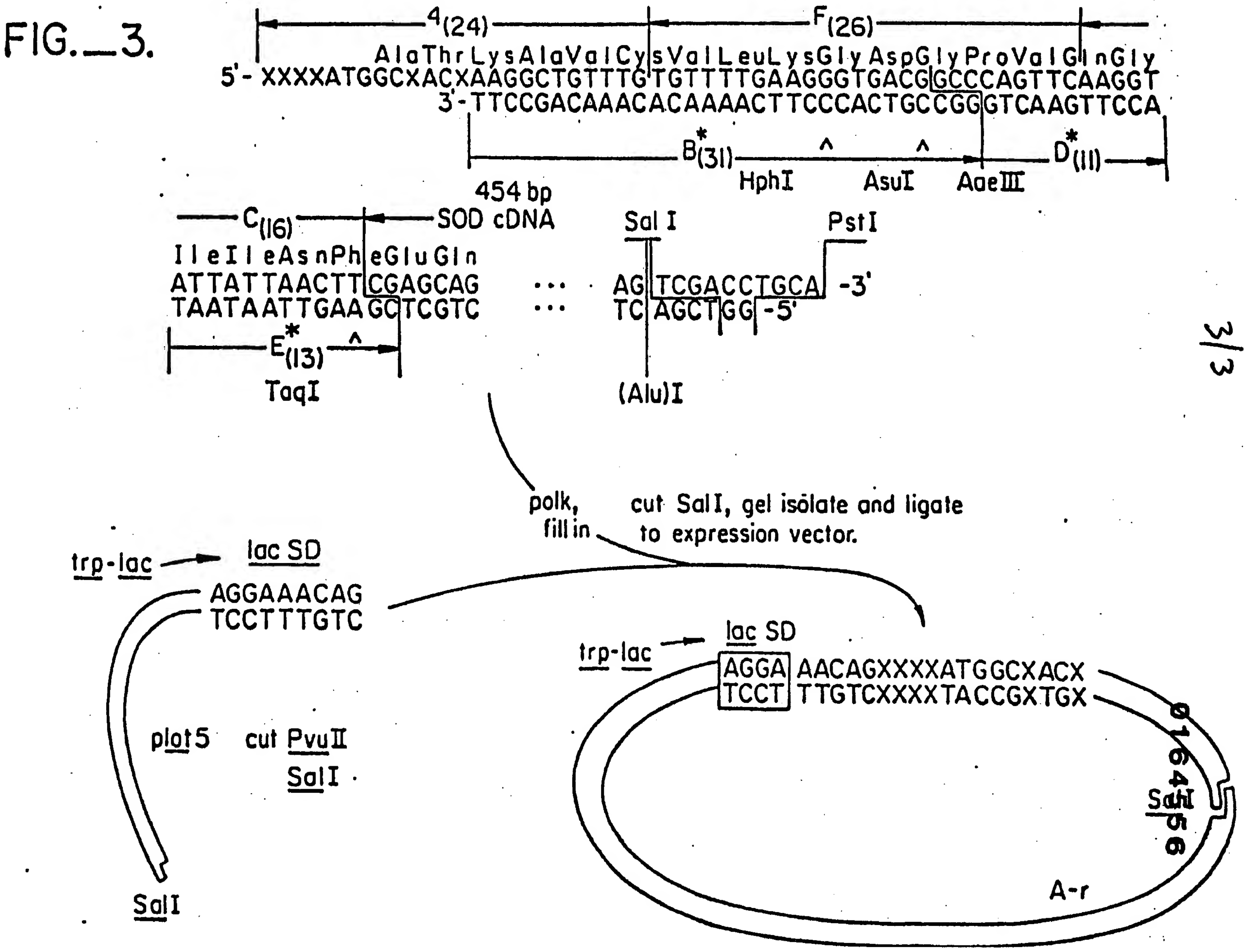


FIG. 2.

FIG._3.


4(24)
F(26)
AlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGly
5'- XXXXATGGCXACXAAGGCTGTTTGTGTTTTGAAGGGTGACGGCCCAGTTCAAGGT
3'-TTCCGACAAACACAAAACTTCCCACTGCCGGGTCAAGTTCCA
B*(31)
D*(11)
HphI
AsuI
AgeIII
454 bp
C(16)
SOD cDNA
IleIleAsnPheGluGln
ATTATTAACTTCGAGCAG
TAATAATTGAAGCTCGTC
E*(13)
TaqI
Sal I
PstI
AGTCGACCTGCA -3'
TCAGCTGG -5'
(Alu)I
polk, fill in
cut SalI, gel isolate and ligate to expression vector.
trp-lac
lac SD
AGGAAACAG
TCCTTTGTC
plot5
cut PvuII
SalI
SalI
lac SD
trp-lac
AGGA
TCCT
AACAGXXXXATGGCXACX
TTGTCXXXXTACCGXTGX
Ø164
Sal
56
A-r